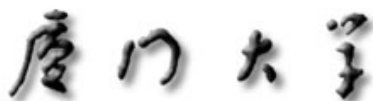


学校编码: 10384

分类号_____密级_____

学号: 24320061152660

UDC_____



硕 士 学 位 论 文

基于 XML 的核磁共振数据管理模型的
研究及应用

Research and Application on
XML-based Data Management Model for NMR

张太彪

指导教师姓名: 曾 文 华 教授

专 业 名 称: 计算机软件与理论

论文提交日期: 2009 年 5 月

论文答辩时间: 2009 年 6 月

学位授予日期: 2009 年 6 月

答辩委员会主席: _____

评 阅 人: _____

2009 年 5 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（
）课题（组）
的研究成果，获得（
）课题（组）经费或实验室的
资助，在（
）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

可扩展标记语言（eXtensible Markup Language, XML）集数据表示、存储、传输和处理为一体，为不同应用程序之间的数据交换提供了一种通用的格式，在各领域得到了广泛的应用。当前，越来越多行业已开始使用 XML 语言来定义各自的新标准。

核磁共振（Nuclear Magnetic Resonance, NMR）自 1946 年首次被观测以来，经过半个多世纪的发展，已经被广泛应用于物理、化学、生物、医学和地学等诸多领域。然而，各主流 NMR 谱仪生产商对 NMR 数据的表述格式不尽相同，数据共享困难，扩展性较差。因此，就发展趋势而言，使用 XML 作为 NMR 数据描述和管理工具是一种很好的选择。

本文主要探讨基于 XML 的 NMR 数据管理模型及其相关应用，为发展拥有自主知识产权的 NMR 谱仪系统软件做准备。论文从核磁共振的基本原理入手，详细分析了 NMR 数据的各种组成部分，介绍了两款主流核磁共振系统软件，并深入剖析了二者的数据描述格式及相应的数据管理模型，指出了当前主流 NMR 数据管理模型的一些不足之处。在此基础上，结合 XML 技术，提出了一种具备扩展性和实用性的基于 XML 的 NMR 数据管理模型；给出了各种 NMR 数据组件的 XML 描述方案，以及基于 XML 的 NMR 数据所适用的存储技术、文档解析技术、XML 数据压缩技术和 XML 数字签名技术。同时，基于此模型开发了具有自主知识产权的核磁共振系统软件，对模型的相关应用进行了探讨。通过对该软件的有关性能分析，验证了本文提出的基于 XML 的 NMR 数据管理模型在处理具体问题时的有效性。

关键词：核磁共振（NMR）；可扩展标记语言（XML）；数据管理模型

Abstract

XML (eXtensible Markup Language), a integrated tool for data collection, storage, transmission and processing, provides a common data structure for the data exchange between various applications, and has been widely used in various fields. At present, more and more industries have begun to use XML to define their own new standards.

After more than half a century's development since it was observed, NMR (Nuclear Magnetic Resonance) has been widely applied to the areas of physical, chemical, biological, medical and earth science etc. However, the mainstream NMR spectroscopy manufacturers have their own standard and there are a lot of difficulties such as non-unified data formats, difficult data sharing and low readability etc. Therefore on the development trend, the use of XML as a NMR data description and management tools is a good choice.

This thesis focused on the NMR-based XML data management model and its applications, aimed at the preparation for the development of the NMR spectrometer system software with intellectual property rights. Starting from the basic principles of NMR, various components of the NMR data are analyzed in detail. Two mainstream NMR system soft wares are introduced. Their data formats and the corresponding data management models are deeply analyzed. And the inadequacies of the current mainstream NMR data management model are pointed out. Then a new scalable and practical NMR data management model is constructed based on the XML The XML description of the NMR data component, storage technology, document analysis technology, XML data compression technology and XML digital signature technology are all given. At the same time, magnetic resonance system software with independent intellectual property rights is developed based on the model and the applications of the model are discussed. The validity of the data management model in dealing with specific questions is verified through related software performance analysis. And the further research direction is determined.

Keywords: NMR; XML; Data Management Model

厦门大学博硕士论文摘要库

目录

第一章 绪论	1
1.1 论文研究的背景及意义	1
1.2 论文的主要研究内容	3
第二章 主流NMR软件及其数据管理模型分析	5
2.1 NMR简介	5
2.2 NMR数据的组成	5
2.2.1 FID	6
2.2.2 磁振频谱	7
2.2.3 脉冲序列	7
2.2.4 参数	9
2.2.5 各种标注	9
2.3 主流NMR系统软件简介及其数据管理模型分析	9
2.3.1 VnmrJ	9
2.3.2 TopSpin	12
2.4 主流NMR系统数据管理模型的不足	15
2.5 本章小结	16
第三章 XML及相关技术分析	17
3.1 XML的语法和结构	17
3.1.1 XML的简明语法	17
3.1.2 XML的文档结构	18
3.1.3 XML验证	18
3.2 XML存储技术	18
3.2.1 文件系统存储方式	18
3.2.2 关系数据库存储方式	19
3.2.3 面向对象数据库存储方式	20
3.2.4 Native XML数据库存储方式	20
3.3 XML解析技术	21
3.3.1 DOM	21
3.3.2 SAX	22
3.3.3 VTD-XML	23
3.3.4 三种解析器的性能对比	24
3.4 XML压缩技术	25
3.5 XML数字签名	26
3.5.1 XML签名简介	27
3.5.2 XML签名语法	27
3.5.3 XML签名的类型	28
3.6 本章小结	29

第四章 基于XML的NMR数据管理模型	30
4.1 有关术语介绍	30
4.1.1 工作区和工作区参数.....	30
4.1.2 页面集合和页面集合参数.....	30
4.1.3 页和页面参数.....	30
4.1.4 层集合和层集合参数.....	30
4.1.5 层和层参数.....	30
4.2 各种NMR数据的XML表示	31
4.2.1 基于XML的FID和谱数据	31
4.2.2 基于XML的参数数据.....	33
4.2.3 脉冲序列的XML表示.....	38
4.2.4 各种标注的XML表示.....	38
4.2.5 NMR数据的文档结构定义	41
4.2.6 NMR数据的存储	41
4.3 基于XML的NMR数据解析与处理	42
4.4 NMR数据安全.....	42
4.5 NMR数据压缩.....	42
4.6 本章小结	43
第五章 基于XML的NMR数据管理模型的应用探讨	44
5.1 NMR系统软件的分析与设计.....	44
5.1.1 NMR系统软件需求分析	44
5.1.2 NMR系统软件的开发环境	46
5.2 基于XML的数据管理模型在NMR系统软件中的实现	46
5.2.1 NMR数据的结构定义及存储	46
5.2.2 NMR数据的解析处理实现	47
5.2.3 NMR数据压缩的具体实现	48
5.2.4 NMR数据的XML签名实现.....	49
5.3 NMR系统软件性能及运行结果	50
5.4 本章小结	54
第六章 总结及展望	55
6.1 总结	55
6.2 进一步的工作和展望	56
参考文献	57
攻读硕士期间的研究成果	60
致谢.....	61

Contents

Chapter 1 Introduction.....	1
1.1 Background and Significance	1
1.2 Outlines of Thesis	3
Chapter 2 Mainstream Data Management Model for NMR.....	5
2.1 Introduction of NMR	5
2.2 Data Component of NMR.....	5
2.2.1 FID	6
2.2.2 Spectrum	7
2.2.3 Pulse Sequences	7
2.2.4 Parameters.....	9
2.2.5 Marks of Spectrum.....	9
2.3 Mainstream Data Management Model for NMR	9
2.3.1 VnmrJ	9
2.3.2 TopSpin	12
2.4 Deficiencies of the Existing Data Management Model.....	15
2.5 Summary	16
Chapter 3 Analysis of XML and Related Technologies	17
3.1 XML Syntax and Structure	17
3.1.1 XML Syntax.....	17
3.1.2 XML Structure	18
3.1.3 XML Validation	18
3.2 Storage Technologies of XML.....	18
3.2.1 File System Approach	18
3.2.2 Relational Database Approach.....	19
3.2.3 Object-oriented Database Approach	20
3.2.4 Native XML Database Approach	20
3.3 XML Analysis.....	21
3.3.1 DOM	21
3.3.2 SAX.....	22
3.3.3 VTD-XML	23
3.3.4 Performance Comparison.....	24
3.4 Compression Technologies of XML	25
3.5 Digital Signature of XML	26
3.5.1 XMLSignature Profile	27
3.5.2 XML Signature Syntax	27
3.5.3 Types of XML Signature.....	28
3.6 Summary	29

Chapter 4 XML-based Data Management Model for NMR	30
4.1 Terms of the Model.....	30
4.1.1 Workspace and Workspace Parameters.....	30
4.1.2 Pages and Pages Parameters	30
4.1.3 Page and Page Parameters	30
4.1.4 Layers and Layers Parameters	30
4.1.5 Layer and Layer Parameters	30
4.2 XML Expression of NMR Data	31
4.2.1 XML-based FID and Spectrum.....	31
4.2.2 XML-based Parameters	33
4.2.3 XML-based Pulse Sequences.....	38
4.2.4 XML-based Marks	38
4.2.5 DTD of the NMR Data	41
4.2.6 Storage of the NMR Data	41
4.3 Analysis of the XML-based NMR Data	42
4.4 Security of the XML-based NMR Data	42
4.5 Compression of the XML-based NMR Data	42
4.6 Summary	43
Chapter 5 Application of XML-based Data Management Model.....	44
5.1 Analysis and Design of the NMR Software	44
5.1.1 Needs Analysis of the NMR Software	44
5.1.2 Development Environment of the NMR Software	46
5.2 Realization of Data Management Model for NMR	46
5.2.1 Structure Definition and Storage for NMR.....	46
5.2.2 Realization of NMR Data Analysis.....	47
5.2.3 Realization of NMR Data Compression	48
5.2.4 Realization of XML Signature	49
5.3 Conclusion.....	50
5.4 Summary	54
Chapter 6 Conclusion and Discussion.....	55
6.1 Conclusion.....	55
6.2 Discussion	56
References	57
Publications	60
Acknowledgements	61

第一章 绪论

1.1 论文研究的背景及意义

核磁共振自 1946 年首次被观测以来^[1], 经过半个多世纪的发展, 已经被广泛应用于物理、化学、生物、医学和地学等诸多领域, 具有了无法取代的优越性。在生物学方面, 可以说 NMR 对于现代生物学的重要问题——蛋白质、核酸在溶液中的三维空间结构的建立是独一无二的工具^[2]。在医学方面, 正是由于核磁共振成像具有非损伤、快速、高分辨率、多参数的优点才使现代医学特别是脑科学出现了许多有深远影响的新发现。在美国的人类基因组计划和人类脑科学计划中, NMR 也同样发挥着不可或缺的作用。2002 年的诺贝尔化学奖和 2003 年的诺贝尔生理或医学奖, 都授予与 NMR 相关的研究, 更是彰显了 NMR 的重要性。近年来更让 NMR 工作者振奋的是 NMR 量子计算机设想^[3], 这一全新的设想将会使未来计算机世界发生翻天覆地的变换。

以此同时, NMR 谱仪技术也得到飞速的发展。由于超导技术的发展, 1964 年出现了 200MHz 超导连续波 NMR 谱仪, 70 年代初已经能生产出 360MHz 以上的 NMR 谱仪, 80 年代已经有了 400MHz 和 600MHz 的 NMR 谱仪。目前, 950MHz 的 NMR 谱仪也已经商业化。

随着经济的发展, NMR 波谱仪在自然科学和人类生活等各领域发挥着越来越重要的作用, 其研究和生产受到了发达国家的高度重视, 年产值更是达到了数十亿美元。改革开放以来, 我国每年引进 NMR 谱仪的数量稳步上升, 近两年的谱仪年进口数量已经超过了 40 台, 预计到 2010 年谱仪的年需求量将达到 80~100 台。可是, 我国的 NMR 波谱仪至今仍完全依赖进口。

现代高分辨率 NMR 波谱仪综合了当代 NMR 波谱学、射频电子学、计算机及自动控制、低温超导和精密加工等领域最新技术, 主要由主控制台、主计算机(包含系统控制与数据处理软件)和超导磁体(包含探头)等部分组成, 研制难度大、费用高。目前世界上仅有两家 NMR 谱仪主要生产厂(Bruker 和 Varian)。由于 NMR 技术对其他领域的发展意义重大, 我国对此也非常重视。为了推进我国 NMR 技术的研发, 由中国科学院武汉物理和数学研究所与厦门大学共同承担的课题“300~500MHz 核磁共振波谱仪的研制”被列入了“十一五”国家科技支撑计划的重点项目之一。该课题旨在自主创新研发 NMR 波谱仪核心技术, 发

展 NMR 应用软件，系统集成 NMR 波谱仪，建立国家 NMR 波谱仪研制平台，提高和增强 NMR 及相关领域的持续创新能力和国际竞争力，打破我国 NMR 波谱仪至今仍完全依赖进口的局面。

该课题的主要技术难点除了仪器硬件外，就是核磁共振系统软件（以下简称“NMR 系统软件”）。作为贵重仪器的系统软件，NMR 系统软件具有专业性强和复杂度高的特点。它不仅具有复杂多样的实验设计和丰富多变的实验方法，也涉及从一维到多维的 NMR 实验的复杂数据处理和显示等操作，这些都对 NMR 系统软件的设计提出了很高的要求。总结起来，该系统软件的任务就是利用计算机对各种 NMR 数据的管理和使用，这其中包括有谱仪控制、NMR 实验设计和执行、一二维数据处理和分析等任务。因此，如何搭建有效的 NMR 数据管理模型，成为了 NMR 系统软件设计的关键。

当前，NMR 谱仪随机软件主要有美国 Varian 公司的 VnmrJ 系统和德国 Bruker 公司的 TopSpin 系统。它们对 NMR 数据均有自己独特的格式和管理模型，而且发展也相对成熟。这两种模型的相似之处在于均采用二进制和普通文本文件相结合的形式来组织和管理各种 NMR 数据。其中，采用二进制格式描述 FID（Free Induction Decay）数据，采用普通的文本格式描述参数文件，采用程序设计语言描述脉冲系列。数据格式的不统一，造成了数据共享相对困难，不仅可读性差，也容易增加实验操作的复杂性。

另一方面，XML 的发展对几乎所有的专业领域都具有深远的意义。如今，许多行业都开始使用 XML 语言来定义新的行业标准。比如，在会计方面，XML 已经发展成为了扩展性商业报告语言 XBRL^[4]。目前已有越来越多的公司把 XBRL 作为财务报告的统一标准。由于 XBRL 在所有的软件格式和输出通道之间提供了更大的互用性，从而使财务信息更加一致，更易理解，而且无需公司改变它们公布财务情况的方式。在电子商务方面，美国许多大公司，如 Microsoft、IBM、SUN、Oracle 等在电子商务文书交换标准化过程中走在了世界的前列，它们与联合国所属的有关机构，如 UN/CEFACT 等达成了协议，共同开发第二代 e-Business 的电子文书标准 ebXML^[5]。ebXML 是基于 XML 标准的，其目标是“提供一种全球化的、开放的、基于 XML 的架构，使不同企业规模的用户之间的电子商务交换能够成功、安全和可靠地进行”。在这种背景下，我国也致力于发展具有自主知识

产权的电子商务交易标准语言cnXML^[6]及其支撑平台。目标是基于XML技术,建立一个符合中国大陆商业习惯、传统和商业流程的B2B、B2C电子商务语言规范,提供一套统一、灵活、开放和可扩充的交易语言。使各个贸易方,包括购买者、供货者、组装者及中介机构等,能够方便地通过电子网络进行各种商业活动,降低企业的运营成本,提高企业产品的竞争力和对市场的响应速度,同时能够方便地与国内外其他电子商务交易语言进行交互。在办公文档方面,选择XML及其相关标准作为定义办公软件文档格式标准语言是当前文档格式定义的技术趋势。现在比较有影响的办公软件文档XML国际标准有两个,一个是与SUN StarOffice相关的OpenOffice^[7]标准,另一个是微软Office Open XML^[8]标准。2008年11月4日,由中电标协文档库技术标准工作委员会制定的电子文档读写接口标准UOML^[9](Unstructured Operation Markup Language),正式被OASIS组织批准为国际标准,成为了中国软件产业领域的第一个国际标准。此外,在高级数据库搜索、网上银行、医药、法律、软件工程等其他领域,XML同样被广泛应用。虽然将XML运用到所有领域需要更长的时间,但这一趋势不可避免并将不断壮大。一些国家甚至对此进行立法,其数据将只能以XML标准进行储存和传递。紧跟这一趋势变得十分重要,因为任何新的应用都将会与XML相适应。因此,就发展趋势而言,使用XML作为NMR数据描述和管理工具是一种很好的选择。

基于以上分析,本文将重点研究基于XML的NMR数据管理模型。开展此项研究主要有以下两方面意义:

第一,推动XML技术在NMR领域的应用,提供一种解决NMR数据描述和管理的XML方案,为制定基于XML的NMR数据管理标准做准备。

第二,克服传统NMR数据格式不统一、共享困难、可读性差等不足,建立一个统一、与平台无关、扩展性良好的全新NMR数据管理模型,为开发国产NMR系统软件,推动我国核磁共振技术深入研究与应用奠定扎实的基础。

1.2 论文的主要研究内容

本论文的工作是围绕NMR数据管理模型的研究而展开的,本文提出了基于XML的NMR数据管理模型,并在此基础上开展了一些应用研究,论文的主要内容安排如下:

第一章,从NMR、NMR谱仪和XML三方面介绍本文的选题背景和研究意义。

第二章，简要介绍了 NMR 的基本原理和相关概念，深入剖析了 NMR 的数据组成和各种 NMR 数据的特点。在此基础上，对 VnmrJ 和 TopSpin 两大主流 NMR 软件及其数据管理模型做了较为详细的分析，指出了它们在数据描述与数据共享方面的不足。

第三章，介绍了 XML 的语法、结构和文档的有效性规则，分析了与论文相关的几个 XML 核心技术，如 XML 存储技术、XML 解析技术、XML 压缩技术和 XML 数字签名技术。为第四章讨论基于 XML 的 NMR 数据管理模型做好准备。

第四章，提出了基于 XML 的 NMR 数据管理模型，包括对脉冲序列、FID、谱、参数和各种图形标注等数据组件的 XML 表示与存储，以及 NMR 数据的 XML 解析处理、XML 压缩和数字签名。

第五章，基于数据管理模型设计完成了具有自主知识产权的 NMR 系统软件。主要介绍 NMR 系统软件的总体需求，并给出了该软件的总体设计，对 NMR 软件中模型的具体实现做了详细探讨，通过对软件有关性能进行分析，验证了该模型的有效性。

第六章，全文总结与展望，提出了今后可以继续展开的研究工作，并就论文存在的不足提出了进一步研究的方向。

第二章 主流 NMR 软件及其数据管理模型分析

2.1 NMR 简介

NMR是一项用来分析化学成分（主要是有机化合物）结构的技术，典型的化合物由碳、氢和氧原子组成^[10]。

最简单的情况下，一个 NMR 实验由三个步骤构成：

- （1）将样品放置在静磁场中；
- （2）使用射频脉冲激发核子；
- （3）检测样品发射出来的信号频率。

具体的激发与响应过程如图 2.1 所示：

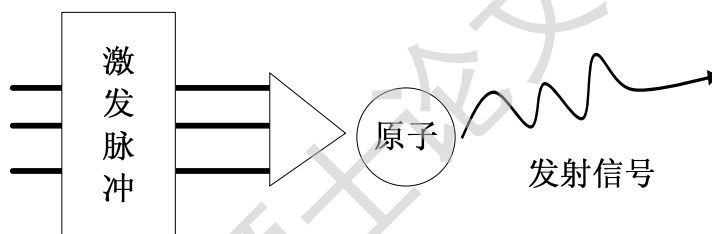


图 2.1 激发与响应

通过分析样品发射信号的频率，可以得到样品中关于原子结合和排列的信息。NMR 激发样品时，核子在不同的频率（称为“共振频率”）产生共振。这些频率就是核子被外来射频脉冲激发而发射出来的频率。而 NMR 信号一般会被绘制成波谱，用于分析频率和强度之间的关系。其中，频率变化的定性信息取决于原子核所处的原子环境，信号的积分强度代表了对信号强度的测量，通过对信号峰下的区域进行积分取得。

2.2 NMR 数据的组成

样品中被激发的原子发射出的信号被谱仪接受并进行傅里叶变换^[11]。接受 NMR 信号的过程被称为采样，即采集数据。一般说来，当执行一个采样时，样品中被激发的原子发射出的信号被采集，此信号被称为 FID。由于 FID 是时间域信号，要对它进行有用的分析必须通过傅里叶变换将其转换成对应的频率域信号，而 FID 和谱便是我们通常所说的狭义上的 NMR 数据。其实，从仪器控制、实

验设计、采样到后期的数据处理、分析都伴随着大量的参数数据，这些参数是整个实验过程的纽带，其管理和控制十分重要。此外，为了保证实验的有效进行，还涉及到众多脉冲系列的设计，在后期的实验结果分析过程中还可能包含有许许多多的标注，如直线、矩形、圆、文本等等。因此，参数、脉冲系列和标注从广义的角度来说都应该称为NMR数据，本论文中研究的正是广义上的NMR数据管理。

2.2.1 FID

FID是核磁共振与核磁共振成像^[12]中最简单的信号形式。受激发的核子对磁共振频谱仪或磁振成像扫描仪的射频线圈造成感应电流而产生信号，并且因发生弛缓而使信号强度逐渐衰减至零，如图 2.2 所示，这种逐渐衰减的信号即称为FID。

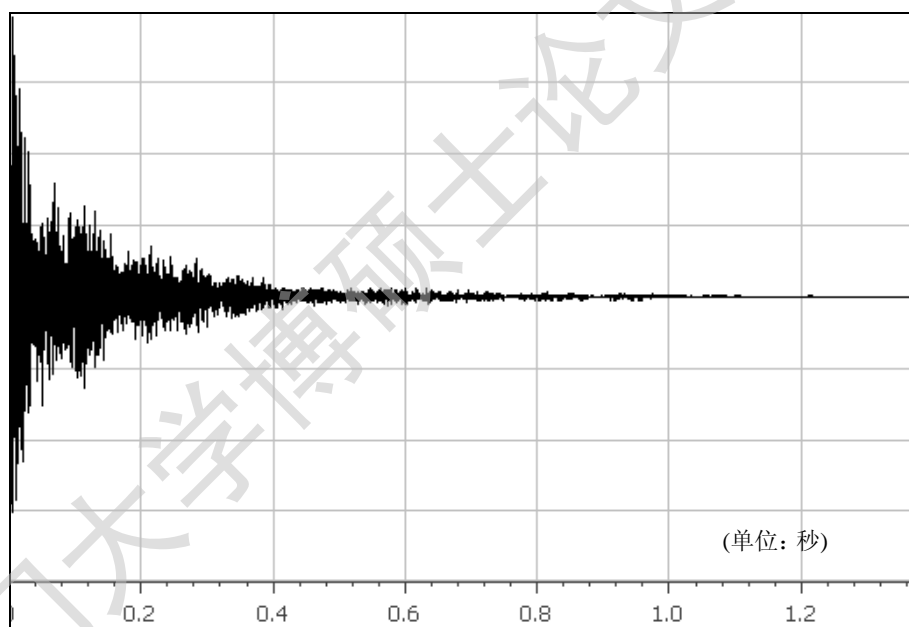


图 2.2 FID 信号

FID信号在多数液体情形下，整个信号形式是振幅呈指数衰减的振荡信号，信号频率为该原子核在如此强度的主磁场下所具有的共振频率——称为拉莫频率^[13]。信号的振荡反映了磁向量在垂直主磁场方向（称为横向）的平面上进动（旋进），衰减则反映了横向上的弛豫^[14]现象。在固体情形，衰减函数则变得复杂，成为高斯函数、洛仑兹函数与正弦函数的混合。

FID信号在某核种被激发后就开始会有信号，然而最前的一段却不能收取信

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库